DOCUMENT RESUME

ED 048 507 VT 012 878

AUTHOR Rosen, Doris B.
TITLE Employment Testing And Minority Groups, As Reported in the Literature. Key Issues Series No. 6.
INSTITUTION State Univ. of New York, Ithaca. School of Industrial and Labor Relations at Cornell Univ.
PUB DATE Jun 70
NOTE 35p.
AVAILABLE FROM Public Information Center, School of Industrial and Labor Relations, Cornell University, Ithaca, New York 14850 ($2.00, Checks made payable to N.Y.S. School of Industrial and Labor Relations)

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Culture Free Tests, Employment Problems, Equal Opportunities (Jobs), *Minority Groups, Task Performance, *Test Bias, *Testing Problems, *Test Interpretation, Test Reliability, Test Validity

ABSTRACT

This report reviews the employment testing of minorities and discusses the limitations of the test results. Statistical validation of test results is stressed as a means to avoid cultural bias and avoid discriminatory screening tests. Noting that fair hiring practices and valid tests may not be enough to provide equal employment opportunities for blacks, the report concludes that educational programs must provide the upgrading necessary for skilled employment. (BH)

employment testing and minority groups

DORIS B. ROSEN

KEY ISSUES: background reports on current topics and trends in labor-management relations—NUMBER 6

KEY ISSUES, a series of reports on current labor-management topics, is a part of the expanded public information service of the New York State School of Industrial and Labor Relations. It includes summaries and surveys of research publications, legislative documents, conference proceedings, and other published materials available in the comprehensive library collection of the school. The reports in this series are intended to provide background information on subjects of interest to management, labor, and the public.

The following reports have appeared to date:

1. *Sensitivity Training: A Review of the Controversy* by Mary Ann Coghill, December 1967, 26 pp.

2. *The Lie Detector in Employment: An Examination of Some of the Problems* by Mary Ann Coghill, July 1968, 24 pp.

3. *Job Satisfaction: A Review of the Literature* by Bonnie Carroll, February 1969, 35 pp.

4. *Portable Pensions: A Review of the Issues* by Jozetta H. Srb, June 1969, 37 pp.

5. *Coalition Bargaining* by Philip J. Schwarz, January 1970, 43 pp.

6. *Employment Testing and Minority Groups, As Reported in the Literature* by Doris B. Rosen, June 1970, 36 pp.

Each report is $2.00 a copy and checks should be made payable to the N.Y.S. School of Industrial and Labor Relations. Orders for individual reports or standing orders for the series may be sent to:

Public Information Center
School of Industrial and Labor Relations
Cornell University
Ithaca, New York 14850

Series Editor: J. Gormly Miller
Associate Editors: Mary Ann Coghill – Jozetta H. Srb

KEY ISSUES SERIES - NO. 6

EMPLOYMENT TESTING AND MINORITY GROUPS, As Reported in the Literature

by

Doris B. Rosen

The purpose of this report is to review the issues related to testing and minority-group employment with special attention to the findings of selected, recent studies carried on within the industrial setting.

June 1970

New York State School of Industrial and Labor Relations
A Statutory College of the State University
Cornell University
Ithaca, New York

ACKNOWLEDGEMENTS

The author wishes to express her gratitude for the very helpful editorial criticism made by Jo Srb, J. Gormly Miller, P. Lewin, N. A. Rosen, W. Hodges, and R. Risley. Also, the articles and materials made available by S. Aker, R. Katzell, E. Mesics, N. A. Rosen and F. Ruch were much appreciated.

D.B.R.
May 2, 1970

TABLE OF CONTENTS

INTRODUCTION

Since the passage of the Tower Amendment to the Civil Rights Act of 1964, special concern has focused on questions related to testing members of minority groups in an employment setting. Many of these questions concern the fairness and relevance of testing disadvantaged groups, and are related in that sense to the national problem of assisting the underemployed to move into rewarding and satisfying employment situations. Although good testing practices alone are insufficient for solving these employment problems, tests, properly used, can be helpful adjuncts in selection and placement. Improperly used tests, however, can and do discriminate against minority groups.

The purpose of this report is to review the issues related to testing and minority groups with special attention to the findings of recent, selected research in the industrial setting, as reported in the literature. Most of the controversy about testing centers on blacks and certain measures of group intelligence: tests of vocabulary, verbal reasoning, arithmetical skills and reasoning, and spatial ability. (Two of the more frequently used tests in private industry are The Wonderlic Personnel Test and The Otis Employment Test.) Only limited data are available for other minority groups and such other types of tests as motor dexterity, vision, trade information, vocational interest, clerical skills, personality, and mechanical comprehension. The research cited in this paper reflects these limitations.

The references, which are cited at the end of the report, will furnish the reader with a fairly comprehensive bibliography. Appendix A provides a number of definitions of basic statistical concepts, and Appendix B lists the sources of guidelines and standards that have been developed to help formulate testing programs.

RECENT BACKGROUND

Industry and the Hard-Core Unemployed

Early in 1968, the National Alliance of Businessmen (NAB) organized Jobs in Business Sector (JOBS) to promote nation-wide efforts to provide jobs for the hard-core unemployed. Sundquist reported that on March 15, 1968, NAB announced that 146,000 unemployed (of whom 87,000 remained on the job) had been hired under the JOBS program and that the program operated in 125 cities.[1]

In June 1969, NAB was merged with Plans for Progress to provide a single effort in providing employment opportunities. Plans for Progress, organized in 1961, preceded NAB and represented concerted efforts by industry to promote fair employment and increase Negro personnel. Recent coverage of NAB activities in Labor Policy and Practice reports that over 15,000 companies participate in the program which is committed to finding jobs for 600,000 hard-core unemployed by June 1971.[2]

Articles by Lockwood (1965 and 1966)[3], and Samuels (1968)[4] review the experiences of various Plans-for-Progress companies in hiring minority group members. Samuels' report focuses particularly on the Ford program for the inner city -- a program in which ability tests were waived and men were hired on the spot. In 1968, Mesics prepared a reference memorandum on the current literature about the hard-core unemployed.[5] He annotated forty-two articles on the topics of poverty, getting and holding jobs, learning and retraining problems, and experiences with integrating the hard-core unemployed into work projects or training. Lockheed, Pitney-Bowes, and Eastman Kodak furnish examples of the impressive efforts of private firms to bring the hard-core unemployed into the work force.

Descriptions of individual NAB company practices to promote fair hiring are reported in Labor Policy and Practice, Volume 6: Fair Employment Practices (490: 1-1076). Included in these descriptions are the following examples. For beginning clerical pool jobs, a Milwaukee subsidiary of the Inland Steel Company eliminated the high school diploma, prior work experience requirements, and all aptitude tests -- keeping only a minimal typing skill test requirement. Another company stopped using three 5-minute tests to select key punch operators when research showed that test scores were not related to job success. In one instance, scores on the test were found to be negatively related to job success: the lower the applicant scored the better were his chances of good performance.

Some writers have voiced skepticism about NAB's progress. Sundquist comments that voluntary organizations often lose their vitality and that NAB's efforts may not be far reaching enough to solve the nation's problems. Samuels reported that only 25 of 400 firms replied positively to a plea from the President to help in hiring or training 500,000 hard-core unemployed. Hayes surveyed 100 of the largest U. S. corporations with their headquarters in New York City

and noted that blacks constituted only 2.6% of the staff.[6] He contends that opportunities for minority groups have not opened up in many jobs or in certain industries. He further argues that employment testing often helps exclude minority group members who could provide badly needed skills in the technical and managerial job areas.

A survey of "Company Experience with Negro Employment" was conducted among 47 companies between July 1964 and January 1966.[7] The experience in these companies, 60 percent of whom were members of Plans for Progress, indicates that a gap does exist between company policy and practice. It appears that blacks were still being hired mainly for the lower skilled jobs and that most companies were unwilling to lower or alter employment standards for blacks. The case history material from the survey suggests that many firms used fairly aggressive recruiting practices during that period but fewer of them engaged in training programs. A report by members of the Personnel Policies Forum (Bureau of National Affairs, 1965) supports these findings.[8]

Finally, Rosen, Goodwin, and Graev's survey data from personnel managers in New York State also revealed interesting relationships between testing and other employment practices.[9] Personnel departments using validated tests -- tests shown to measure what they purport to measure -- as opposed to those using non-validated or no tests, were characterized by greater commitment to testing but were less flexible in their attitudes toward making special allowances for culturally disadvantaged applicants. A considerably higher percentage of those using validated tests indicated that their organizations had special programs for hiring disadvantaged individuals.

Employment Tests and Fair Employment Rulings

The Motorola Case. The issue of fairness in testing was brought to national attention in July 1963 by the Motorola Case when Leon Myart filed a complaint of alleged discrimination with the Illinois Fair Employment Practices Commission.[10] Previously, Myart, a Negro, had applied for a job as "analyzer and phaser" at Motorola. He had been interviewed, given a five-minute intelligence test, and sent home without being told whether or not he would be selected. After two weeks, he filed a complaint stating that he had passed the test and was rejected because of his race.

The case was heard before an examiner in January 1964. Although Motorola was unable to produce the specific test in question, they maintained that Myart had failed it. (There are indications, however, that Myart was capable of attaining a passing score, and that reliability and validity data for the test presented by the defendants was inadequate.) The examiner directed that Myart be given employment and the test suspended because it was obsolete and had been normed on "advantaged groups" and did not "lend itself to equal opportunity to qualify for the hitherto culturally deprived and disadvantaged groups."

In November 1964, a commission supported the examiner's findings, and in April 1965, a Circuit Court decision upheld the commission's findings. But, in March 1966, the Illinois Supreme Court reversed the judgment on the basis that the evidence did not support the allegation of unfair employment practice.[11] Despite the ruling of the Illinois Supreme Court, the Motorola case became especially important because it contained the first fairly well-publicized legal recognition that a particular test might be inappropriate for use with disadvantaged groups.

Government Action. In enacting Title VII of the Civil Rights Act of 1964, Congress made it unlawful for employers of 25 individuals or more, labor unions, and employment agencies to discriminate against an individual because of race, color, religion, sex, or national origin. The Tower Amendment to Title VII of the Civil Rights Act of 1964, Section 703(h), permitted employers to give and act only upon the results of any professionally developed ability test which was not designed, intended, or used to discriminate against the above listed individual groups.

The Equal Employment Opportunity Commission (EEOC) was established in 1964 to implement the above law and assist in the elimination of employment discrimination by investigating complaints and violations of the 1964 Civil Rights Act. In cases where allegations of discrimination are found to be soundly based, provision for conciliation is made. The conciliatory procedure is voluntary, and unresolved cases may be referred to the Justice Department for further action.[12] Approximately 15,000 charges were filed with the Commission in 1968, and according to Cooper and Sobel, 15-20% of all the charges filed under Title VII involved a testing issue.[13]

In 1965, the Department of Labor instituted Executive Order 11246 to eliminate discrimination by government contractors and sub-contractors. The Office of Federal Contract Compliance (OFCC) bears the responsibility for administering government policy and has issued regulations concerning the obligations of contractors to provide equal employment opportunities. Failure to comply can result in sanctions which delay or cancel the awarding of contracts.

Both the EEOC and the OFCC have issued guidelines for employment testing which are based upon professionally determined standards for test selection, administration, and validation. Basically, the intent of both sets of guidelines is to promote fair testing practices, but some differences of emphasis may be noted. One major difference between the EEOC and OFCC guidelines is that the former apply to all types of positions while the latter currently exclude professional, technical, and managerial jobs. The EEOC requires evidence of test validity for all occupational levels, but only in situations where a selection test is producing a high rate of rejections among minority group members. The OFCC order requires contractors to establish test validity in blue collar and clerical jobs, regardless of the rejection rate for minority groups. The order applies to contractors and sub-contractors having contracts of $10,000 or more and employing more than 1,000 employees.

Psychologists too have been active in setting up guidelines, publishing research, and generally reformulating their own positions. The American Psychological Association (APA) standards on testing have been generally adopted by the EEOC and OFCC. To review and discuss relevant issues on this topic, the APA recently (September 1969) held a workshop on "Approaches to Compliance with Governmental Regulations on Fair Employment."[14] Many excellent materials that provide guidelines and standards have been developed over the past five years to help formulate testing programs. A partial list of references to them is given in Appendix B at the end of this report.

During very nearly the same time these guidelines have emerged, researchers and practitioners in the field of employment testing have become more aware that the increasing use of tests makes it imperative to evaluate the whole subject of test usage. The expanding literature during the past decade covers such topics as the misuse of personality tests, over-dependence on testing, and the questionable validity of some of the tests used for selection and placement in private industry. And within these areas there is a considerable body of research on the reasons for and the effects of the differential test performances of racial groups.

ISSUES IN FAIR EMPLOYMENT TESTING

It has been suspected that in some instances tests have been used to deliberately screen out blacks. The concern of this report, however, is with unintentional rather than intentional discrimination. Most commonly blacks have been rejected for employment because they have failed to meet certain test standards.

Differential Test Performance of Blacks and Whites

Studies of performance on ability tests in a variety of settings generally indicated lower test scores for blacks than whites. Extensive reviews of the literature on differential test performance of racial groups by Dreger and Miller, Himmelstein, Jensen, Klineberg, and Shuey debate the issue of the effects of nature vs. nurture on intellectual performance.

Before reviewing the conclusions of these authors, however, it is important to note that in the distribution of intelligence test scores for blacks and whites, we find overlap in the range of population scores. Thus, although the average test score may be lower for blacks than whites -- and for low vs. high socioeconomic groups -- there are whites who score lower than most blacks and blacks who score higher than most whites. In short, the distribution of intelligence test scores in the population is continuous for all groups rather than being distinctly separate, i.e., all blacks are not in the low sample nor are all whites in the high sample.

Shuey has summarized a large number of studies made over the past fifty years on testing and Negro intelligence in various age, occupational, educational, and societal groups.[15] Her overall conclusion -- which is disputed by most psychologists -- is that the consistently lower performance by Negroes points to the presence of "native" differences between Negroes and whites. More recently, Jensen caused considerable controversy by concluding, after examining the results of both his own research and that of others, that heritability of intelligence is quite high.[16] He states the position that the low distribution of tested IQ for Negroes is not necessarily accounted for by environmental factors but is primarily attributable to biological factors involved in the types of abilities which are differentially distributed in the population as a function of race and social class. He further states that compensatory education programs have failed because they are predicated on reducing the environmental gap rather than focusing on the specific skills that deprived children are capable of learning.

Many other psychologists reach the opposite conclusion from the same body of research evidence. Dreger and Miller favor the environmental deficiency theory to explain lower Negro test performance.[17] But they comment on the limited number of meaningful comparative studies in which complex experimental procedures and designs are used. Klineberg states that there is no

scientifically acceptable evidence for the view that ethnic groups differ in innate abilities.[18] He also comments that "this is not the same as saying there are no ethnic differences in such abilities." He favors an environmental explanation of differences. Miner, Campbell, and Roberts point out that Negro-white differences appear to be decreasing.[19]

The discussion of nature vs. nurture may have very little practical value for the employment setting where the major question is: "Can the prospective employee do the job?" The implications of the discussion, however, are of value in understanding the relationship of test scores to ability, in any setting. A comprehensive statement on race and intelligence, formulated by the Society for the Psychological Study of Social Issues Council, reflected an unequivocal stand that is important to this issue.[20] The six major points espoused were:

1. Comparable cultural and educational background for whites and Negroes reduces markedly the difference in intelligence test scores.

2. Large numbers of black people lack social, economic, and educational opportunities available to most whites.

3. Compensatory education's failures are due mainly to such factors as inadequate planning, size, and scope.

4. Posing questions in the simplistic terms of nature vs. nurture ignores the essence and nature of human development and behavior.

5. Intelligence tests tend to be biased against blacks and while such tests can predict school achievement, they are not accurate measures of innate ability.

6. To prove genetic differences is most difficult, especially since the common criterion for race is usually based upon skin color.

Cultural Bias in Testing

If minority groups are denied widespread exposure to adequate educational, social, and economic advantages, it is commonly agreed that test performance may be negatively affected, or depressed. Thus, a test can be culturally biased when it measures possible verbal, quantitative, or spatial skills to which a minority group may have had little exposure. The crux of the problem is related to Krug's comment that absolute measures of achievement will generally underestimate the potential of an individual from a sub-standard environment.[21]

Anastasi has written extensively on the topic of cultural bias.[22] In her discussion of testing the disadvantaged, she contends that:

> "...if we rule out cultural differentials from a test, we might thereby lower its validity against the criterion we are trying to predict. The same cultural differences that impair an individual's test are likely to handicap him in school work, job performance, or any other activity we are trying to predict."

In fact, Krug describes attempts to minimize the cultural differences between groups by developing "culture-free" or "culture-fair" tests. These tests have been shown to lack practical significance. Ash notes that culture-fair tests do not seem to measure aptitudes or characteristics significantly related to such ordinary measures of performance as job tenure, production, or foreman's ratings.[23]

Anastasi further notes that in order to predict such future behavior as job performance, tests need to be highly relevant to work output, supervisory ratings, or other specific measures of job performance. And finally, she comments that the characteristics of test scores are less likely to vary among cultural groups when the test is intrinsically relevant to job performance rather than to non-job related factors.

Psychometric Definition of Test Bias

In discussing cultural bias, it is also important to consider the problem in its psychometric context. Test bias, psychometrically defined, relates specifically to over-prediction or under-prediction of job criterion measures. Thus, if a test consistently under-predicts performance on the job for a given ethnic or socioeconomic group, it shows bias against this group.

Cleary discusses two sources of psychometric bias.[24] First, she defines item bias in which average scores on particular items in a test differ markedly for different groups. Thus, if most whites answer an item correctly while most blacks answer the same item incorrectly, this would suggest item bias against blacks.

Second, test bias is indicated if a predicted level of performance on some criterion is consistently too high or too low for members of the sub-group. For example, if it is found that high scores on a given clerical test are associated with great accuracy for clerk-typists, while low test scores are associated with inaccuracy, and, the association holds equally for all groups, the clerical test would be an unbiased predictor of clerical accuracy. However, if it is found that a white sample consistently scores higher than a black sample on a pre-employment test but the two groups do not differ in their actual ability to perform on the job, the test would be considered biased. In practice, then, the high scoring white applicant might be hired in preference to the low-scoring black applicant.

To summarize briefly, since minority groups have a history of poor test performance, it is axiomatic that their rejection rate in employment situations has been higher than for whites. In many instances, as indicated in the above discussion, the "purpose" of the tests has not been intentionally to bar blacks from employment, but in effect this has been the result. There seems to be general agreement that the "screening out" procedure is not unfair so long as the rejectees are incapable of satisfactory job performance. However, considerable evidence indicates that many employment tests are unintentionally biased because they have not been satisfactorily validated. In other words, tests often have been used without being properly related to tasks involved in the job and with little or no evidence of their statistical validity.

Lack of Validity in Employment Testing

Some of the improper testing practices frequently encountered among employers are listed in the EEOC guidelines on testing. These problems generally reflect a lack of well-conceived validation procedures:

1. Testing programs which have been developed without adequate professional advice and which have not been based upon careful job analysis procedures.

2. Use of arbitrary cut-off scores without firm evidence that these scores differentiate between successful and unsuccessful job performance.

3. Insufficient records on employee performance which make it impossible to conduct test validation studies.

4. Use of semi-secret devices, such as personality tests, which are difficult to validate locally.[25]

Recent research furnishes evidence that industrial validation is not commonly practiced. Ash reports a survey by the University of Wisconsin's Industrial Relations Center in which only seven percent of 152 companies reported that all their tests had been validated locally against on-the-job performance measures.[26] Nearly 60 percent of the companies had not validated any of their tests.

A survey of personnel directors in New York State by Rosen, Goodwin, and Graev revealed that 35 percent of 107 respondent organizations used validated tests, while 49 percent used non-validated tests, and 16 percent used no tests at all.[27] These percentages are even less impressive when it is noted that the original questionnaire elicited responses from only 33 percent of the sample to whom it was mailed. In a study of three major local governmental units in the greater Miami area, Rosen and Serino found virtually no empirical evidence of test validity, despite the extensive use of testing for selection purposes by these same agencies.[28]

Cooper and Sobel cite a court case brought under Title VII of the Civil Rights Act of 1964 in which company applicants for all jobs, including janitor, were required to pass a test battery.[29] The employer attempted to show validity by using a sample of only the top eight percent of employees rather than a proper sample that included all levels of employees.

Many excellent position statements concerning the need for test validation have been made during the past few years. The most concise and forceful one was written by Guion:

> "The principal implications of Title VII, then, seem clear enough.... People using employment tests had better gather data to demonstrate that their tests are valid as predictors of relevant aspects of job behavior for all classes of applicants, and if these tests are found invalid, they ought not be used."[30]

Procedures for Test Validation

The question of how validity is determined in the employment situation is crucial to any consideration of testing and bias. In many cases, it is considered highly desirable to establish empirical validity; that is, to determine statistically, by a correlation coefficient, the degree of relationship between test performance and a relevant criterion measure. However, in some instances, validity can be established on the basis of "experts" checking to see if the test content appears to reflect accurately the purpose of the test. For example, a test of mechanical comprehension should contain items pertaining to recognized areas of mechanical knowledge.

Concurrent validation is a frequently used statistical approach in which the degree of relationship is determined by a correlation coefficient between the test scores of applicants and job performance measures of a currently employed sample. This relationship is then generalized to the applicant population. Another type of validation, predictive (or longitudinal), determines the correlation coefficient of test scores of the applicant population and some criteria of their own performance at a later date. This latter method is not often used because of the problems of following up employees over a period of time.

Bennett points out some of the limitations of studies using concurrent validity procedures.[31] Two of these limitations are: range restriction of talent in the employee population and unreliable criterion measures. For example, the best employees in the current group may have already been promoted and the poorest may have been terminated; thereby restricting the range of talent in the employed group. Supervisory ratings may be overly generous or harsh depending on the particular raters involved; thus providing unreliable criterion measures. In this connection, Bennett feels that "trainability" can be used as a reliable and meaningful criterion in test validation because it reflects both prior knowledge

and an index of willingness and ability to learn. Training programs can be especially valuable procedures for helping to select the potentially satisfactory employees among individuals previously considered unemployable because of low test scores.

Many psychologists have stressed the importance of preventing range restriction in the employee sample by initially using a low enough cut-off score to include low-scoring individuals who might prove to be successful employees. However, all employees should be followed up over time to determine in what way the raising or lowering of such cut-off scores relates to successful job performance.

Content validation, validating a test rationally on the basis of how closely it reflects actual job requirements, is sometimes used in instances when predictive or concurrent validation is unfeasible; for example, this procedure may be used with samples which are too small to establish a statistical validity coefficient. Ruch states that job analysis as a method of content validation is "acceptable" and can be highly useful in smaller organizations when the "work forces are not large enough to support the use of more elegant statistical designs."[32]

An excellent illustration of content validation is reported by Aker.[33] At Olin Mathieson, series of tests for use in specific plants have been developed for craft occupations. For example, the machinist's test consists of three parts: Trade Information, Blue Print Reading, and Shop Calculation. The content is specifically related to work performed by the maintenance department at one particular site in Ohio. In the test manual, the authors of the test series caution that in cases where content validation is used, no attempt is made to establish a validity coefficient or to predict future level of performance. The test score represents a level of knowledge or achievement in a job related area, and score distribution is limited to "pass" or "fail" categories, the dividing line between the two being arbitrarily determined.

Moderator Techniques in Research Design

Formerly, it was considered appropriate to lump various sub-groups together (i.e. whites and blacks) to establish validity coefficients for tests. The procedure has been subject to some criticism recently because predictive validity can vary for different groups. A currently used technique to investigate the problem of potential bias is a statistical design using moderator variables.

According to Saunders[34], the moderator variable design provides a method for studying situations where membership in one or another distinct group (like race) may "influence" the relationship between two variables, such as intelligence and productivity. In other words, the researcher is trying to find out whether the relationship of test scores to job related criteria differs for blacks and whites. If test performance relates differently to job performance for the

white group than for the black group, it may indicate the need to use different prediction (or regression) equations and to develop separate job performance expectancy tables for each group. In this way, accuracy of prediction for members of both groups can be maximized by moderating, or controlling, the effects of race. In cases where investigation shows no differences between ethnic groups, in terms of the way test performance relates to criterion performance, the use of separate prediction schemes is not indicated.

Although there is general support for using moderator techniques with different sub-groups, several reservations have been expressed about how to use them. First, this technique requires a large enough sample of minority group members to permit adequate groupings for research purposes. Second, groups must be homogeneous and distinct from each other. Polarmo points out in addition that moderators on which subjects are categorized may not be truly homogeneous.[35] Ghiselli and Sanders caution also against assuming that a given test is measuring all individuals or groups with the same degree of reliability, and they urge researchers to check for reliability before using a moderator variable prediction scheme.[36]

RECENT RESEARCH ON TESTING AND CULTURAL FAIRNESS IN THE INDUSTRIAL SETTING

Reviews of the literature on testing and bias in industry reveal a dearth of empirical research in this area.[37] This paucity of material is the result of the nature of recent federal legislation as well as the complications involved in conducting research in the industrial setting. However, research evidence is becoming increasingly available. For example, fourteen studies which relate to validity and fairness in testing were summarized by the 17th Annual Workshop in Industrial Psychology.[38] These studies are important from a methodological as well as a substantive point of view and illustrate the kinds of data now emerging on the question of racial differences in the relationship of test scores to job-performance criteria.

Office Workers

Ruda and Albright studied racial differences in scores on selection instruments and related these differences to the subsequent job performance of 327 hired applicants in a large office.[39] Correlational techniques were used in which the predictors -- a weighted application form, and the Wonderlic Personnel Test -- were related to job performance criteria and termination rates. The major findings of this study can be summarized as follows.

1. High scores on the weighted application blank were associated with a tendency to remain on the job for both racial groups.

2. High scores on the Wonderlic Personnel Test for whites were associated with a tendency to leave the job.

3. No relationship between Wonderlic scores and termination was found for blacks. (As a group, blacks tended to score lower and stay on the job longer than whites.)

4. No relationship was found between either the Wonderlic or weighted application blank scores and job performance criteria of promotion and engineering time standards.

A major conclusion from this study was that the Wonderlic test was being used incorrectly for selecting whites and was irrelevant for Negroes.

Machine Shop Trainees

Using race and socioeconomic status as moderators, Tenopyr made two studies of machine shop trainee selection tests.[40] In the first study, 500 applicants for machine shop trainee jobs were given verbal, numerical, and

space visualization aptitude tests. Data showed that whites consistently outperformed blacks even when socioeconomic status was controlled (or moderated). Low performance by blacks on the space visualization tests leads the author to conclude that "...the Negro job applicant may be at as great a disadvantage when so-called 'culture-fair,' spatial tests are used in selection, as when verbal tests are utilized." Moore, et al. supported this conclusion in a recent study of ethnic differences with an industrial selection test battery.[41]

In the second study by Tenopyr, 167 trainees, 84 whites and 83 blacks, were selected on the basis of a composite score of the above tests. The relationship between the three tests and ten training achievement criteria were studied by a method of correlational analyses. The major findings in the second study were the following. First, there were no significant differences in the relationship of test scores to level of achievement between high and low socioeconomic groups. Second, there were significant differences between black and white groups in the relationships of test scores and achievement measures. In six out of ten analyses, putting black test scores into prediction equations based solely upon white test scores would have resulted in "over-prediction" of average black performance. Thus, the discrimination would favor blacks. The author notes that there is some reason to believe that these particular findings may have been the result of criterion bias in the raters' judgements.

The author concludes that there is need for further investigation before a decision is made to use different prediction equations or passing scores for whites and blacks. She also urges further study in the areas of cultural differences and achievement motivation.

Toll Collectors (N. Y. Port Authority)

Lopez described procedures by the New York Port Authority to select employees for the new position of Female Toll Collector.[42] One hundred and eighty-two collectors -- 102 black and 80 white -- were appointed from 2,000 applicants on the basis of test scores and interview ratings. These selection instruments were validated later against four criteria: absence rate, tolls accuracy, continued employment, and supervisory ratings. The major findings in this study were the following.

1. Although blacks achieved lower scores on selection instruments than whites, these differences were not related to lower job performance.

2. Separate correlational analyses for racial groups revealed dissimilar patterns of association. For example, for blacks, high scores on predictors were unfavorably associated with attendance and continued employment, but were favorably associated with tolls accuracy. For whites, high scores on predictors were unfavorably associated with tolls accuracy and continued employment, but only a high score on the written test was associated unfavorably with attendance. Also, high test scores were

significantly related to supervisory ratings for white workers, but not for black workers. For the latter group, the interview rating was significantly associated with supervisor ratings.

Lopez argues for using an overall assessment strategy and the proper weighting of instruments in terms of their differential predictive validity. However, the results of this study have been seriously questioned because validity coefficients were tested for statistical significance after being raised by correcting for restriction of range in the sample. It is not considered acceptable procedure to apply significance tests to validity coefficients that have been adjusted in this way.

Other Employment Situations

An extensive series of five investigaions on testing and fair employment -- supported by Ford Foundation funds -- was made by Kirkpatrick et al. over a two year period.[43] Five different employment situations were studied in which a total of 1,208 job incumbents -- 795 white, 325 black, and 88 Spanish -- were employed. Comparisons were made between test score means and job performance on various criterion measures for different ethnic groups in comparable jobs. The influence of race and cultural status was moderated in an attempt to improve predictability for different ethnic, racial, and socioeconomic groups.

A summary of these studies is presented on the chart following on p. 16.

Other industrial research cited in the summary report of the 17th annual workshop in Industrial Psychology showed a somewhat similar pattern of findings. Gordon found minimum qualifying scores for technical schools on the Airman Classification Battery to be equally valid for whites and blacks.[44] Data from a study by Mitchell and others revealed that the Wonderlic Personnel Test and biographical data were not valid predictors of performance criteria for either white or black semi-skilled plant workers.[45] Grant and Bray found ability and mechanical and dexterity tests to be equally valid predictors of training success for both white and black telephone and installation repairmen.[46] And data by Maslow indicate that tests from French's Kit of Selected Tests for Reference Aptitudes and Achievement Factors were valid predictors of supervisory ratings and a job knowledge test for medical technicians from both racial groups.[47]

Evidence of some over prediction for blacks emerged in the latter two studies; that is, the level of performance was not as high as test scores would indicate. Explanations for this phenomenon will be explored in the summary.

Conclusions From Research in the Industrial Setting

The conclusions drawn by Kirkpatrick, et al., in the summary of their studies on Testing and Fair Employment are applicable to the current body of research:[48]

	STUDY I	STUDY II	STUDY III	STUDY IV	STUDY V
SUBJECTS	136 female clerks in an insurance company: 34 Negroes, 102 whites.	71 female clerks in same insurance company as Study I: 33 Negro, 38 white.	Male trainees, low socioeconomic backgrounds, for general maintenance program. Group I: 53 Spanish, 53 Negro, 31 white. Group II: 35 Spanish, 44 Negro, 46 white.	Students in a professional nursing program. Group I: (St. Louis)--31 Negro, 101 white. Group II: (D.C.)--32 Negro, 39 white.	Female clerks from various insurance companies (pooled): 98 Negro, 437 white.
TESTS	Short Employment Test, interviewer ratings.	Battery of vocabulary, numerical checking, and coding subtests; also 2 non-verbal reasoning tests.	Gates Reading Survey and Numerical Ability Test of DAT -- used for guidance, not selection.	Pre-nursing and Guidance Examinations.	Clerical selection test with various sub-tests.
CRITERIA	Varied termination, merit rating, pay job grade.	Supervisory estimates on a rating form.	Graduation vs. termination performance and proficiency test scores.	State licensing exam scores, classroom performance ratings, training program termination time.	Salary level, supervisory ratings.
VALIDITY	Tests not valid, no significant correlation with criteria.	Effectiveness rating best predicted by numerical test for whites, coding test for Negroes. Non-verbal tests not significantly related to criterion.	Different patterns for each group. No one test valid for all groups.	Different for each group; tests more generally valid for whites than for Negroes.	Coefficients for two groups nearly identical. Used concurrent validation.
MODERATOR VARIABLES	Validity not improved by using social-economic status as moderator variable.	Not useful, whether based on race or cultural status.	Cultural status as moderator did not improve validity. Race as moderator improved prediction criterion for Negro-white sample.	Cultural status as moderator did not improve prediction. Race as moderator improved validity.	Cultural status and race not useful in improving prediction.
FAIRNESS	No significant difference between test or criteria means for racial groups.	No significant evidence of unfairness. Possibility of one instance of rater bias against whites.	Some under-estimation of minority group criteria scores.	Performance under-predicted for Negroes. Because whites and Negroes went to different schools, the finding is questionable.	No conclusions since subjects were from different companies and rated by different supervisors.

Summary of the Findings of Kirkpatrick, et al.[43]

1. Tests can differ in validity and degree of validity for different ethnic groups. In some instances, tests may be valid for whites and not blacks, but in other cases the reverse may be true. Thus, test validity or lack of validity is not necessarily generalizable from one group or setting to another group or setting.

2. Tests can discriminate unfairly between ethnic groups. Minority group criterion performance is most commonly under-estimated by tests.

3. The moderated prediction technique can be useful in improving the validity of tests for different ethnic samples. In some cases, including race as a moderator variable improved the correlation between test score and job performance measures.

4. A cultural status index derived from standard biographical data is not likely to be useful in improving the correlation between test score and criterion performance.

5. Non-verbal tests are not necessarily fairer for minority groups than are verbal tests.

6. Job training appears to improve scores on some types of selection tests for all groups, not just minority groups.

OTHER RELATED RESEARCH FINDINGS

Because this report focuses on testing and minority groups in the industrial setting, there has been no attempt to include a large body of literature which relates to minority groups in public service. The reader who is interested in this aspect of the subject will find the bibliographies published by the Civil Service Commission extremely valuable.[49] Related research from other non-industrial settings illustrates some of the situational and motivational factors which are particularly relevant to evaluating test performance by minority groups.

Test Practice Effects

Tests are sometimes criticized as being racially biased because blacks have had only limited opportunity to practice test taking skills. In 1950, Hay described a hort test designed to serve as a practice instrument to overcome nervousness and inexperience.[50] The author maintained that such a device improved testee's attitudes. More recently, other research has focused upon investigating the importance of test practice on subsequent test performance.

Dubin and others investigated the hypotheses that extra test practice, extra testing time, or both would increase mental ability test performance for some groups of high school students more than for others, (i.e. for blacks over whites and low over high social-economic group members.)[51] However, results showed all groups profitted to a comparable extent when the practice and extra time procedures were followed. The authors conclude that the testing procedure itself is not discriminatory.

Droege investigated the long-range effects of practice on the General Aptitude Test Battery (GATB) used by the United States Employment Service (USES).[52] Significant practice effects occurred for all aptitudes for sub-samples. The score increases were retained in varying degrees for all aptitudes even after three years, and these gains appeared not to be related to such variables as age, years of education, or aptitudes. Droege comments that re-testing appears unnecessary unless a person has been exposed to training and experience which might increase his knowledge and test performance.

In other research by the United States Employment Service, Droege and Bemis reported on the development of a non-reading edition of the GATB which correlated (.75) with a reading measure for a literate sample of 471 people.[53] This test appeared also to differentiate between the abilities of educationally deficient individuals and retarded individuals. Dvorak, Droege and Seeler describe other aspects of the USES program to assist the under-employed: determining applicant's potential ability to take the GATB; and the continuing test validation procedures using performance and training criteria.[54]

The limited data on test practice effects from the above studies indicate that the testing procedure itself does not necessarily discriminate against blacks. In specific instances, irrelevant factors related to a particular examiner, inadequate testing facilities, or problems of the testee may contribute to low test performance. In these cases, retesting may be desirable. In general, however, retesting of applicants does not appear to be routinely necessary unless some learning experience has occurred between tests. Where individuals have a low level of literacy, appropriate tests like the non-reading edition of the GATB may need to be used to assess ability.

Educational Factors

Studies from educational settings provide further insight into academic performance, current recruitment problems, and certain psychological aspects of adjustment by blacks in desegregated facilities.

Discussion of research by Cleary, Cleary and Hilton, Campbell, Munday and Stanley and Porter has revealed that in general standardized ability tests appear to be useful and relatively unbiased predictors of academic success for Negroes.[55] However, the findings that some academic ability tests are not biased and that test scores seem to relate to academic success for both blacks and whites does little to solve the problems of recruitment by industry.

In discussing the problem of black students from southern colleges, Holland points out that blacks are often critically low on standard achievement exams.[56] Dugan's data on comparative test performance between white graduates of northern colleges and Negroes from southern colleges shows less than 10% of Negroes as opposed to 50% of the whites qualified for employment consideration.[57] He also reports other data collected by recruiters at a large electronics corporation indicated 60 college graduates were contacted to every hire among Negroes in southern colleges as opposed to 15 such contacts in northern integrated colleges. The major reason for rejection of Negroes was the inability to meet employment standards. It should be noted that most of the data on which this finding is based was for blacks from Negro colleges in the South where academic performance has traditionally been low for both blacks and whites.

Both Holland and Dugan, however, note instances where minority group members performed better than test scores indicated. Therefore, the task for industry, as aptly stated by Holland, is to identify the well-prepared or potentially competent youth although he may lack certain cultural background factors and may be deficient in certain courses of study because of the limitations of the college.

And finally, Katz has written a review of the evidence of educational desegregation on the scholastic achievement of blacks.[58] He notes conditions which are detrimental to black performance, among these are social rejection, fear of competition with whites, inadequacy of previous training, and unrealistic inferiority feelings.

The result of Katz's findings with bi-racial team experiments on the college level indicates more passivity and social compliance among blacks than whites -- probably caused by social threat or fear of failure. Blacks performed better under white examiners in task situations which were characterized by low social threat or failure. However, when severe threat was introduced, they performed far more adequately under a black examiner. The authors whose works were reviewed by Katz, thus, infer that the perceived probability of attaining a white standard of success was an important determinant in the high motivation for blacks performing tasks with the white examiner. Blacks also appeared freer to exhibit hostile expressions with black than with white testers.

Data presented by Katz have practical implications for understanding the problems which sometimes occur with black employees. For example, fear of competition, rejection, and failure can lead to underperformance and conflict in the work situation. The data also indicates that the tendency of some blacks to form cliques and to be poorly motivated can be viewed psychologically as adaptive behavior to reduce tension in an environment which is perceived as threatening and hostile.

IN SUMMARY

The Motorola case brought the topic of testing and "cultural bias" to the attention of employers, and the public, in the middle 1960's. And, with the passage of the Tower Amendment to Title VII of the Civil Rights Act, it became imperative for industrial organizations to review their testing programs and practices for relevancy and fairness. The EEOC and OFCC were established subsequently to implement the Act. Both agencies have issued guidelines to promote good testing practices to help eliminate discrimination in employment.

It has been suspected that in some instances tests have been used deliberately to screen out blacks. But most often, blacks have been rejected for employment because they have failed to meet certain test standards. They have often been in the position of being passed over for employment because of low test scores which most psychologists agree are related to inferior educational and social opportunities. Recent research indicates that the gap between the test scores of whites and blacks appears to be closing.

Test Validation

In general, industrial leaders have expressed the view that test standards should not be lowered for certain individuals or groups. But the major controversy centers around the evidence that test standards have not always been relevant to, or related to, job performance measures. In short, surveys indicate that test validation has not been commonly practiced. Thus, in many cases, individuals or groups seeking employment have been unfairly rejected because many employment tests have not measured what they purport to measure.

The most important question in the industrial setting is not "how high is Joe's test score?" but "how does Joe's test score relate to job performance?" The answer to the latter question is determined by using proper validation techniques. Tests are commonly validated in three ways: concurrently, predictively, or by content analysis. The first two methods measure the relationship of test performance and job performance statistically, by using a correlation coefficient. When using either of these validation procedures, it is important to include the entire range of performers in order to determine if a test can properly discriminate between successful and unsuccessful employees. Content validation utilizes the approach of rationally determining if test material corresponds to job requirements. This method is most suitable where small samples preclude using statistical procedures.

Newer strategies in research design involve moderated prediction techniques to determine whether the relationship (expressed by a correlation coefficient) of test scores to performance ratings differs for minority groups. Thus, the "influence" of race and cultural differences is assessed in terms of how the effects of these variables on the relationship between the test score and performance measures can be moderated or controlled.

Research Summary

A review of the empirical literature on testing and employment bias indicates that test validity and fairness cannot be inferred but must be statistically or rationally determined for specific tests, or test batteries, with different ethnic groups and job situations, in relation to particular job performance criteria. The purpose for which the test is to be used must also be considered.

Tests found to be invalid serve no useful purpose and may contribute to discriminatory employment practices. Where evidence reveals differential validity for racial or ethnic groups, appropriate job expectancy measures may need to be computed separately for different groups to increase test fairness. In other cases, it may be appropriate to use different tests for blacks and whites. However, the data from a small number of studies indicate that non-verbal measures such as those involving spatial concepts may not be more "culture-fair" than are the more traditional verbal tests. Nor do many of these tests appear to be practical in the employment setting.

The subject of over-prediction of criterion performance -- i.e., consistently lower job performance ratings than are predicted by test scores -- needs further exploration. It is important to determine whether low criterion performance is related to bias in the criterion, such as unreliable rating procedures, or to other intervening factors, such as poor employee motivation.

The research also indicates that more studies of a longitudinal nature are needed. It is important to follow employee work behavior over a period of time if we are to understand the complexities in the work environment which influence and temper the relationship between an applicant's test score and his job performance. The research by Katz and others illustrates some of the situational and motivational factors that need to be fully analyzed. Problems of poor motivation, lack of communication, employee withdrawal, and the forming of black cliques all need to be understood in terms of adjustmental behavior by blacks to cope with an environment that they perceive as alien or even hostile.

CONCLUSIONS

Community, business, and industrial organizations have undertaken the task of promoting fair hiring practices and providing jobs for the hard-core unemployed. There are indications that these voluntary activities by industry may not be far reaching enough to solve the problems. Neither will good testing practices alone solve the pressing problem of upgrading black skills in a continually advancing technological society. Blacks must also be prepared educationally to exit from the dwindling unskilled labor force into the clerical, managerial, and professional fields which provide the expanding job opportunities in the 1970's. The basic research described in this report is relevant to solving intelligently the immediate problem of assisting large numbers of the underemployed to move into more rewarding and satisfying jobs. The notion that the major function of tests is to screen out undesirables is obsolete. Properly-validated and fairly administered tests are extremely useful adjuncts to the efforts to provide jobs, and equal employment opportunities, for blacks.

APPENDIX A

DEFINITIONS OF TERMS

Correlation Coefficient - a statistic which indicates the relationship between two phenomena. For example, a correlation might express the degree of association between two test scores, or between a test score and supervisory rating. Correlation coefficients can range from minus 1.0 to plus 1.0 indicating on the extremes a perfect negative or perfect positive relationship. A 0.0 correlation indicates no relationship.

Criterion - simply the measure or score that a test is trying to predict such as academic grades or job performance. Often times it is difficult to reach agreement on what constitutes good job performance. To the extent that criteria are poorly defined or "biased", the ability of a test to function as a predictor of performance is lessened.

Moderator Variable Technique - A statistical procedure for studying situations where membership in one or another group can influence the relationship between two variables.

Multiple Prediction Scheme - This involves combining mathematically, in an optimal weighting scheme, various information such as test scores, interview ratings, and biographical data in order to predict performance.

Norm - performance of the group or groups on whom a test has been standardized. It is always important to determine if norm groups given for particular tests are generalizable and appropriate to a specific setting.

Reliability - refers to stability or consistency of test scores over a period of time. Reliability is reported by a correlation coefficient. For tests it is desirable to have coefficients ranging in the .80's and above.

Validation Techniques - The process of validating a test involves determining the degree of the relationship (correlation coefficient) between a test score and some measure of performance (criterion score). Tests are often validated concurrently, predictively, or by content (see section under validation for further discussion).

Validity - refers to the accuracy with which a test measures what it intends to measure. Validity is also reported in terms of a correlation coefficient. Validity coefficients for tests range from .30 - .60. Validity coefficients tend to be lower than reliability coefficients partly because of a reflection of the difficulties involved in predicting behavior in a complex situation from a standardized test score.

APPENDIX B

REFERENCES ON GUIDELINES FOR TESTING

1. American Psychological Association, Standards for Education and Psychological Tests and Manuals (Washington: APA, 1966).

2. "Guidelines for Testing Minority Group Children," Journal of Social Issues, April 1964, pp. 129-145.

3. "Job Testing and the Disadvantaged," American Psychologist, July 1969.

4. National Association of Manufacturers, Equal Employment Opportunity Compliance and Affirmative Action (New York: NAM, 1969), pp. 41-56 and 95-97.

5. U. S. Equal Employment Opportunity Commission, "Guidelines on Employment Testing Procedures," 1966.

6. U. S. Office of Federal Contract Compliance, "Validation of Employment Tests by Contractors and Sub-Contractors Subject to the Provisions of Executive Order 11246," 1968.

REFERENCE NOTES

1. James L. Sundquist, "Jobs for the Hard-Core Unemployed," Personnel Administration, September-October 1969, pp. 8-11.

2. For further detail, see: Bureau of National Affairs, Fair Employment Practices, in the series Labor Policy and Practice, vol. 6 (Washington: BNA), pp. 490: 1-41. In the above volume of this series, all of section 490 titled "Selection of Minority Personnel" may be of interest to the reader.

3. Howard C. Lockwood, "Progress in 'Plans for Progress' for Negro Managers," in Selecting and Training Negroes for Managerial Positions, Proceedings of the Executive Study Conference (Princeton, N.J.: Educational Testing Service, 1965), pp. 1-22; and "Critical Problems in Achieving Equal Employment Opportunity," in The Industrial Psychologist: Selection and Equal Employment Opportunity (A Symposium) from Personnel Psychology, Spring 1966, pp. 3-10.

4. Gertrude Samuels, "Help Wanted: The Hard-Core Unemployed," The New York Times Magazine, January 28, 1968, p. 27.

5. Emil A. Mesics, The Hard-Core Unemployed, an Annotated Bibliography (Ithaca: New York State School of Industrial and Labor Relations, Cornell University, 1969), pp. 1-4.

6. Ulric Hayes, Jr., "Equal Job Opportunity: The Credibility Gap," Harvard Business Review, May-June 1968, pp. 113-120.

7. National Industrial Conference Board, "Company Experience with Negro Employment," Studies in Personnel Policy, No. 201 (New York: NICB, 1966), vols. I and II, and "Digest of Studies."

8. Bureau of National Affairs, "The Negro and Title VII," Personnel Policies Forum, Survey No. 77 (Washington: BNA, July 1965), pp. 1-17.

9. Ned A. Rosen, Nina P. Goodwin, and Lawrence G. Graev, "Personnel Testing and Equal Employment Opportunity," Industrial and Labor Relations Research, November 1967, pp. 19-23.

10. For details on the early hearings in the case see: Robert L. French, "The Motorola Case," The Industrial Psychologist, August 1965, pp. 29-50.

11. Motorola Inc. v. Illinois Fair Employment Practice Commission et al., Illinois Supreme Court, No. 39297, March 24, 1966; 61 LRRM 2596.

12. For further detail, see: National Association of Manufacturers, Equal Employment Opportunity Compliance and Affirmative Action (New York: NAM, 1969); pp. 41-56 discuss the most recent guidelines for employment testing. Guidelines adopted by the Commission on August 24, 1966 are listed in Appendix F, pp. 95-97.

13. George Cooper and Richard B. Sobel, "Seniority and Testing under Fair Employment Laws: A General Approach to Objective Criteria of Hiring and Promotion," Harvard Law Review, June 1969, pp. 1598-1679. This article approaches the subject from a legal point of view. The authors have served as counsels for employees in the cases discussed.

14. Brian S. O'Leary, ed., "Approaches to Compliance with Governmental Regulations on Fair Employment," in Section II of the 17th Annual Workshop in Industrial Psychology (Dr. Raymond Katzell, Session Leader). Unpublished.

15. A. M. Shuey, The Testing of Negro Intelligence, 2nd ed. (New York: Social Science Press, 1966).

16. Arthur Jensen, "How Much Can We Boost IQ and Scholastic Achievement?" Harvard Educational Review, Winter 1969, pp. 1-123.

17. Ralph Mason Dreger and Kent S. Miller, "Comparative Psychological Studies of Negroes and Whites in the United States: 1959-65," Psychological Bulletin, September 1968, Part 2, Monograph Supplement, pp. 1-58; see also, "Comparative Psychological Studies of Negroes and Whites in the United States," Psychological Bulletin, September 1960, pp. 361-402.

18. Otto Klineberg, "Negro-White Differences in Intelligence Test Performance: A New Look at an Old Problem," American Psychologist, April 1963, pp. 198-203. See also: Philip Himelstein, "Research with the Stanford-Binet, Form L-M: The First Five Years," Psychological Bulletin, March 1966, pp. 156-164.

19. J. B. Miner, "Constraints on Personnel Decisions: Individual and Cultural Differences," Chapter 2 in Personnel Psychology (New York: The MacMillan Company, 1969), pp. 11-45; Joel T. Campbell, "The Problem of Cultural Bias in Selection: I. Background and Literature," in Selecting and Training Negroes for Managerial Positions, Proceedings of the Executive Study Conference (Princeton, N.J.: Educational Testing Service, 1965) pp. 57-64; and S. O. Roberts, "The Problem of Cultural Bias in Selection: II. Ethnic Background and Test Performance," in Selecting and Training Negroes for Managerial Positions, Proceedings of the Executive Study Conference (Princeton, N.J.: Educational Testing Service, 1965), pp. 65-75.

20. "Motivation and Academic Achievement of Negro Americans," The Journal of Social Issues, Summer 1969. This issue of the publication is devoted entirely to the subject of motivation and academic achievement of American Negroes.

21. Robert E. Krug, "The Problem of Cultural Bias in Selection: III. Possible Solutions to the Problem of Cultural Bias in Tests," in Selecting and Training Negroes for Managerial Positions, Proceedings of the Executive Study Conference (Princeton, N.J.: Educational Testing Service, 1965), pp. 77-90.

22. Anne Anastasi, Psychological Testing, 3rd ed. (New York: MacMillan Company, 1968), pp. 558-565, quote is on p. 558; see other comments by this author in: Invitational Conference on Testing Problems, Testing Problems in Perspective, Anne Anastasi, ed. (Washington: American Council on Education, 1966).

23. Philip Ash, "Race, Employment Tests, and Equal Opportunity," Journal of Intergroup Relations, Autumn 1966, pp. 16-26.

24. Anne T. Cleary, "Test Bias: Prediction of Grades of Negro and White Students in Integrated Colleges," Journal of Educational Measurement, Winter 1968, pp. 115-123. See also: Cleary and T. L. Hilton, "An investigation of Item Bias," Educational and Psychological Measurement, Spring 1968, pp. 61-75.

25. See reference to National Association of Manufacturers, Note 12 above.

26. See Ash, Note 23 above.

27. See Rosen, et al., Note 9 above.

28. Ned A. Rosen and G. Serino, "Employee Selection Practices Used in Three Local Government Units, with Special Emphasis on Tests and Their Likely Impact on the Employment of Black Personnel." **In press.**

29. See Cooper and Sobel, Note 13 above.

30. Robert Guion, "Employment Tests and Discriminatory Hiring," Industrial Relations, February 1966, pp. 20-37; quote is on p. 27.

31. George K. Bennett, "Factors Affecting the Value of Validation Studies," Personnel Psychology, Autumn 1969, pp. 265-269.

32. Floyd L. Ruch, "Critical Notes on 'Seniority and Testing under Fair Employment Laws' by Cooper and Sobel in The Harvard Law Review of June 1969," Unpublished, 1970.

33. Stanley Aker, et al., OMA L Test Series for Crafts (Olin Mathieson Chemical Corp., 1969).

34. D. R. Saunders, "Moderator Variables in Prediction," Educational and Psychological Measurement, Summer 1956, pp. 209-222.

35. Jean M. Palormo, "Test Validation -- a New Must," The Personnel Administrator, November-December 1969, pp. 5-12.

36. E. E. Ghiselli and E. P. Sanders, "Moderating Heteroscedasticity," Educational and Psychological Measurement, Winter 1967, pp. 581-590.

37. See for example: John R. Hinrichs, "Psychology of Men at Work: Employment of Minority Groups," Twenty-First Annual Review of Psychology, 1970, pp. 541-542; William A. Owens and Donald O. Jewell, "Personnel Selection: Fair Employment," Twentieth Annual Review of Psychology, 1969, pp. 420-422; James J. Kirkpatrick, et al., Testing and Fair Employment, (New York: New York Univ. Press, 1968); and Phyllis Wallace, Beverly Kissinger, and Betty Reynolds, "Testing of Minority Group Applicants for Employment" in EEOC Office of Research and Reports, Research Report 1966-67.

38. Studies referred to in the text and cited below as numbers 39-40 and 42-47 are summarized in the summary of Section II of the 17th Annual Workshop in Industrial Psychology by O'Leary, see Note 14 above.

39. E. Ruda and L. E. Albright, "Racial Differences on Selection Instruments Related to Subsequent Job Performance," Personnel Psychology, Spring 1968, pp. 31-41.

40. Mary L. Tenopyr, "Race and Socioeconomic Status as Moderators in Predicting Machine Shop Training Success," presented in a symposium on Selection of Minority and Disadvantaged Personnel. American Psychological Association, Washington, D. C., September 4, 1967.

41. Clay L. Moore, John F. MacNaughton, and Hobart G. Osburn, "Ethnic Differences within an Industrial Selection Battery," Personnel Psychology, Winter 1969, pp. 473-482.

42. Felix M. Lopez, "Current Problems in Test Performance of Job Applicants," from The Industrial Psychologist: Selection and Equal Employment Opportunity (A Symposium) from Personnel Psychology, Spring 1966, pp. 10-18.

43. James J. Kirkpatrick, et al., Testing and Fair Employment (New York: New York University Press, 1968). The chart summarizing the results of the five studies is based mainly on Chapter 2, "Summary of Methods and Results," pp. 13-26.

44. Mary A. Gordon, "A Study of the Applicability of the Same Minimum Qualifying Scores for Technical Schools to White Males, WAF, and Negro Males," San Antonio, Texas: Human Resources Research Center, Lackland AFB. Technical Report 53-54, 1953.

45. M. D. Mitchell, L. E. Albright, and F. D. McMurray, "Biracial Validation of Selection Procedures in a Large Southern Plant," Proceedings of the 76th Annual Convention of the American Psychological Association, 1968, 3, 575-576.

46. D. L. Grant and D. W. Bray, "Validation of Employment Tests for Telephone Company Installation and Repair Occupations," Experimental Publication System, 1969, 1, Ms. No. 021C.

47. A. P. Maslow, Symposium on Test and Job Performance of Negroes and Whites, American Psychological Association, Washington, D. C., September 2, 1969.

48. See Kirkpatrick, et al., Note 43 above.

49. U. S. Civil Service Commission, Equal Opportunity Employment, Personnel Bibliography Series No. 29, 1968; see particularly the section on "Employment Programs for Minority Groups," pp. 29-75: "General" -- pp. 29-40, "Employment Practices - Recruitment, Selection, Testing" -- pp. 41-56, "Professional and Executive Positions" -- pp. 56-61, and "Minority Groups in Federal, State, and Local Governments" -- pp. 62-75.

50. Edward N. Hay, "A Warm-up Test," Personnel Psychology, Summer 1950, pp. 221-223.

51. Jerry A. Dubin, Hobart Osburn, and Darvin M. Winick, "Speed and Practice: Effects on Negro and White Performance," Journal of Applied Psychology, February 1969, pp. 19-23.

52. R. C. Droege, "Effects of Practice on Aptitude Scores," Journal of Applied Psychology, August 1966, pp. 306-310.

53. R. C. Droege and S. E. Bemis, "New Developments in Aptitude Testing of the Educationally Deficient," American Psychologist, August 1964, p. 521.

54. Beatrice J. Dvorak, Robert C. Droege, and Joseph Seiler, "New Directions in U. S. Employment Service Aptitude Test Research," Personnel and Guidance Journal, October 1965, pp. 136-141.

55. See Note 24 for Cleary and Hilton; J. Campbell, "Testing of Culturally Different Groups," Educational Testing Service Research Bulletin RB-64-34 (Princeton, N.J.: Educational Testing Service, June 1964), p. 25; Leo Munday, "Predicting College Grades in Predominantly Negro Colleges," Journal of Educational Measurement, December 1965, pp. 157-159; J. C. Stanley and A. C. Porter, "Correlation of Scholastic Aptitude Test Score with College Grades for Negroes Versus Whites," Journal of Educational Measurement, Winter 1967, pp. 199-218.

56. Jerome H. Holland, "Preparation of the Negro College Graduate for Business," in Selecting and Training Negroes for Managerial Positions, Proceedings of the Executive Study Conference (Princeton, N.J.: Educational Testing Service, 1965), pp. 23-40.

57. Robert D. Dugan, "Current Problems in Test Performance of Job Applicants," in The Industrial Psychologist: Selection and Equal Employment Opportunity (A Symposium) from Personnel Psychology, Spring 1966, pp. 18-24.

58. Irwin Katz, "Review of Evidence Relating to Effects of Desegregation on the Intellectual Performance of Negroes," American Psychologist, June 1964, pp. 381-399.